Application of free energy calculations for drug design

de Beer, S.B.A.

2012

document version
Publisher's PDF, also known as Version of record

CHAPTER 7

NON-SELECTIVE BINDING OF TRIPEPTIDES TO THE OLIGOPEPTIDE BINDING PROTEIN A (OPPA), MEDIATED BY WATER MOLECULES

Adapted from: **Non-selective binding of tripeptides to the oligopeptide binding protein A (OppA), mediated by water molecules.**
Stephanie B. A. de Beer and Chris Oostenbrink
Manuscript to be submitted

Abstract

The periplasmic oligopeptide binding protein A (OppA) represents a well-known example of water-mediated protein-ligand interactions. OppA relies on water molecules to accommodate a broad range of ligands with different physico-chemical properties. Different water configurations have been observed for different oligopeptides, binding to the active site. This observation, combined with the highly flexible peptidic character of the ligands, make the OppA system an excellent test case for advanced computational approaches.

Here we describe free energy calculations for three tripeptides binding to OppA using thermodynamic integration and one-step perturbation approaches. Thermodynamic cycles were constructed for three different peptides binding to OppA.

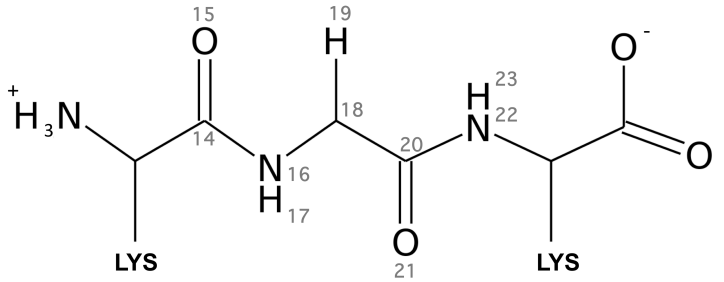
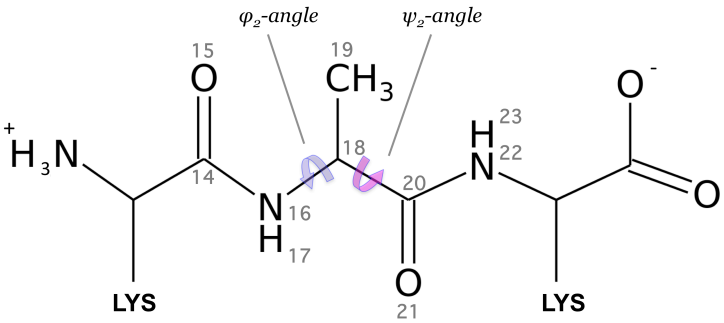
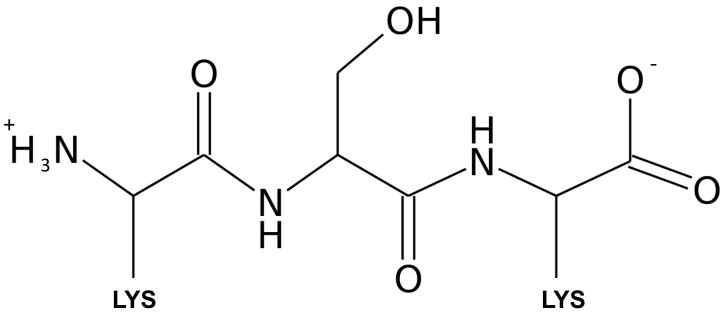
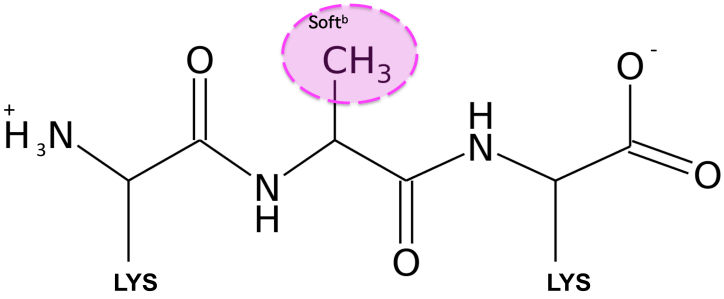
For the tripeptides free in solution, it was observed that the difference in flexibility between an alanine and a glycine residue leads to a slow convergence of the calculations. This can be overcome by either extending the simulations, or by slightly modifying the Hamiltonian of the tripeptides, followed by appropriate corrections to obtain unbiased free energy differences.

Significant discrepancies between independent simulations of the tripeptides in complex with OppA remained, which could be traced to a slow relaxation of the water molecule network within the active site. For the structurally most similar tripeptides, KGK and KAK, the one-step perturbation method offers an elegant solution, leading to a fair agreement of the calculated values for the free energy of binding with the experimental affinities. For the free energy calculations involving larger side chains, significantly longer simulations are most likely required.

7.1 Introduction

In the previous chapter, the role of water molecules mediating protein-ligand recognition and ligand binding was reviewed, as well as the challenges for computational chemists to include them correctly in their studies, at all levels described in Figure 1.3 in Chapter 1. One of the most challenging tasks is to predict not only whether active site water molecules are being replaced or conserved upon ligand binding, but to also quantify the thermodynamic characteristics.[1]
The current chapter addresses the prime challenges regarding the role of water molecules in binding, using the periplasmic oligopeptide binding protein A (OppA) as a case study. OppA, part of the large family of bacterial ABC import systems, plays a crucial role in the transport of peptides across the cell membrane in Gram-negative bacteria, providing them with peptide nutrients.[2]
Towards the end of the 1980s, the unusual character of this protein to selectively bind oligopeptides (rather than single amino acids and dipeptides) was discovered.[3] However, the protein does not seem very selective for specific oligopeptide sequences. It soon became clear that OppA (together with other peptide-binding proteins) plays important roles in microbial cell signaling and virulence, therefore forming an important target for antibacterial drugs, potentially inhibiting bacterial growth for, *e.g.*, *Escherichia coli*, *Salmonella typhimurium* and *Borrelia burgdorferi*.[3b,4] The latter bacterium is the causative agent for Lyme disease. Its genome lacks genes encoding for amino acid synthesis, hence it is dependent on OppA for the uptake of nutrients.[4b] Blocking OppA mediated peptide transport appears to be a useful strategy against Lyme disease.
Another strategy involving OppA as an anti-microbial target in drug design, is based on a 'Trojan horse' principle. By designing peptide-based antibiotics such that they are recognized by OppA, it is highly probable that they will be transported into the bacterial cell and carry out their anti-microbial action from inside the cell.

Overall, OppA is an important drug target, and knowledge of the mechanism of substrate recognition, selectivity and the role of water molecules in this binding process, is of utmost interest to modern day drug design. In this chapter, we explore the binding of a series of tripeptides to OppA from *Salmonella typhimurium*. The binding to this protein by a series of 28 tripeptides is previously characterized experimentally by X-ray crystallography and isothermal titration calorimetry (ITC).[5] One striking observation from these experiments is, that the protein seems very promiscuous towards the middle amino acid in the tripeptides, binding apolar, polar, aromatic and charged amino acids with similar affinity. The structures of the three ligands under consideration here (KGK, KAK and KSK) are shown in Table 7.1, together with the experimentally determined affinities in terms of the Gibbs binding free energy ($\Delta G_{bind,exp}$).

Tripeptide ligand (KXK)	***Structure***	$\Delta G_{bind,exp}$ [a]
Lys–Gly–Lys (KGK)		-33.6
Lys-Ala-Lys (KAK)		-41.1
Lys-Ser-Lys (KSK)		-42.0
Lys-*REF*-Lys (K*R*K)		

[a]Experimental data adapted from reference 5b,c

[b]Soft atom with Van der Waals parameters of a CH3-group and a softness parameter α of 1.51

Table 7.1. Molecular structures of the KXK tripeptides, and the experimentally determined data (adapted from references 5c and 5b for the OppA-KXK tripeptide complexes.

The reference state, used in the one-step perturbation simulations to represent 'real' ligands KAK and KGK is also depicted (K*R*K). Atom numbers and φ_2- and ψ_2-angles for the backbone moiety of the central residue, X, are indicated as well. In KAK, the φ_2-angle around the N_{16} - CA_{18} bond is defined by atoms C_{14}, N_{16}, CA_{18} and C_{20}, the ψ_2-angle around the CA_{18} - C_{20} bond is defined by atoms N_{16}, CA_{18}, C_{20} and N_{22}.

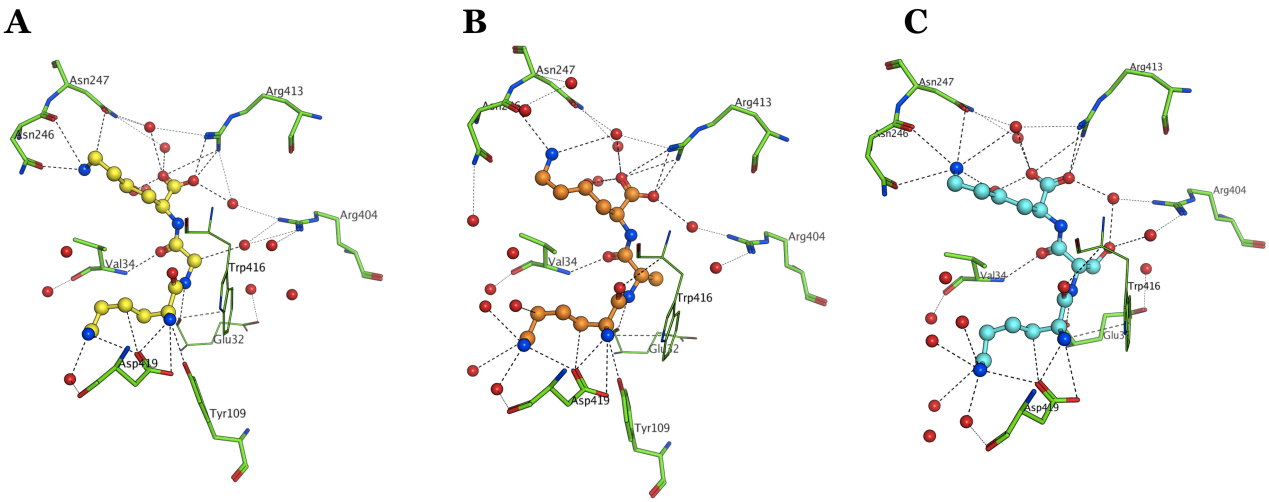


Figure 7.1. Depiction of the binding of the tripeptide ligands in the active sites of the respective crystal structures of OppA in complex with (**A**) KGK (PDB code 1B3L),[5c] (**B**) KAK (PDB code 1JET)[6] and (**C**) KSK (PDB code 1B51).[5c] OppA active site residues which are suggested to be important for ligand recognition and stabilization of the water-network, are labeled and represented in licorice representation. The tri-peptides are shown in ball-and-stick representation, in the colors yellow (KGK), orange (KAK) and cyan (KSK), respectively. All active site water molecules within 0.45 nm of the ligand are shown in red spheres as well. Hydrogen bonds are represented by dotted lines.

These peptides were selected based on their structural similarity, but different binding free energies. The active sites of the three crystal structures of KGK, KAK and KSK bound to OppA are shown in Figure 7.1. All active site water molecules within a range of 0.45 nm of the ligand are depicted as red spheres.

As it appeared from the published crystal structures (Figure 7.1), the observed lack of selectivity for tripeptides is facilitated by a layer of ordered, interfacial water molecules in the active site surrounding the side chains of the bound peptides.[6] These peptidic side-chains make few direct interactions with the protein, while for the backbone moiety there are recurring protein-ligand interactions observable in the complexes (Figure 7.1). Obviously, water molecules play a large role in facilitating the unselective binding of KGK, KAK and KSK, and it is therefore of utmost importance that they are being incorporated in computational modeling studies. However, as addressed in Chapters 1 and 6, computational tools often neglect solvation effects. Referring back to Figure 1.3 in Chapter 1, solvent molecules are rarely included explicitly at Levels 1, 2 and 3, while they are commonly included at Levels 4 and 5. However, great challenges persist for advanced computational studies conducted at these levels as well, since the incorporation of water molecules at this level of computational accuracy also remains a difficult topic,[7] as discussed in Chapter 6.

The unusual promiscuous binding of tripeptides to the OppA protein, and the wealth of published experimental data, provides an excellent test case to study the relationships between structure and thermodynamics and the (mediating) effect of solvent molecules on these.

Applying more rigorous methods on the OppA system, such as free energy calculations which have their origin in statistical thermodynamics, theoretically allows us to properly incorporate the effects of active site water molecules on the binding thermodynamics.
The experimental characterization, *i.e.* the crystallographic and calorimetric information, was published in the mid-90's.[5a,5c] However, to the best of our knowledge, so far no *quantitative* computational study has been reported, which includes the individual contributions of the solvent and the loss of conformational flexibility of the protein and the ligands. Previous attempts involved either empirical correlations based on 3D-QSAR[8] or implicit solvent calculations,[9] or they used a coarse grained model to describe the system, losing the atomic detail of individual water molecules by definition.[10]
It obviously remains a challenge to quantitatively reproduce the experimentally determined binding affinities of the OppA-KXK complexes. In our opinion, a successful calculation on this challenging system can only be performed by a method that explicitly allows for flexibility of the protein and tripeptidic ligands and that explicitly takes the individual interfacial water molecules that fill up the active site into account. Hence, molecular dynamics (MD) simulations together with the underlying statistical mechanics is the method of choice to study tripeptide binding to OppA.

Estimation of differences in free energy of binding from molecular dynamics simulations can be carried out by various approaches, including thermodynamic integration (TI)[11] and free energy perturbation (FEP),[12] which is applied in the one-step perturbation (OSP).[13]
Using these methods, differences between two (or more) thermodynamic states can be estimated using exact statistical thermodynamics from molecular simulations by sampling the relevant conformations, and extracting the appropriate ensemble average. Alchemical transformations are applied to connect the Hamiltonians of two states. In TI, state A is gradually changed into state B in a number of discrete steps, while OSP is a potentially more versatile method, using a (not necessarily physical) reference state which is represented by a single, broad ensemble for which overlap with the ensembles of states A and B is possible and thus free energy differences with respect to a set of real compounds can be deduced. While TI is computationally expensive and can only take a few compounds into account, the OSP is more efficient as it can potentially incorporate more compounds into one screen.
In this chapter, the difference in binding affinity between the tri-peptide ligands, shown in Table 7.1, to OppA is studied, assessing the binding free energy by applying the TI and OSP methods.

7.2 Methods

7.2.1 Thermodynamic integration

In TI, the coupling parameter approach is used to alchemically connect the two Hamiltonians describing two states A and B over a path defined by λ. This scaling parameter is introduced such that at $\lambda = 0$, the Hamiltonian $H(\lambda)$ describes state A, and at $\lambda = 1$, it represents state B. At intermediate λ-values the exact functional form of $H(\lambda)$ is continuous between A and B and represents, possibly unphysical, intermediates.[14]

The free energy difference ΔG_{AB} between the two states in a selected environment (*e.g.* within the protein binding site or free in solution) can be calculated by integration along this path, according to Eq. (7.1),[11]

$$\Delta G_{AB} = G_B - G_A = \int_0^1 \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda \qquad \text{(Eq. 7.1)}$$

The angular brackets denote the ensemble average of the derivative of H with respect to λ, which is obtained from independent simulations at discrete intermediate values of λ. The final free energy estimate is typically obtained by numerical integration over the ensemble averages.

By calculating the free energy differences between two (related) ligands when they are free in solution and when they are bound to the protein and applying the proper thermodynamic cycle (*e.g.* the one in Figure 1.4 in Chapter 1), the relative binding free energy may be calculated.

In this study, we applied TI to calculate the free energy difference between all pairs of the tripeptide ligands, depicted in Table 7.1. In order to monitor convergence of the TI simulations, the free energy difference corresponding to the backward processes was also calculated. The difference between forward and backward calculations is called the hysteresis and should approach zero for converged simulations.

Initial TI calculations showed poor convergence for the ligands free in solution. To circumvent larger sampling times, alternative TI calculations were carried out with a slightly modified Hamiltonian, excluding specific intra-molecular interactions. In a next step, corrections are applied to this TI profile to obtain the correct free energy difference (see Section 7.3.2 for details).

7.2.2 One-step perturbation

If one focuses on the thermodynamic cycle depicted in Figure 1.4 in Chapter 1, the free energy difference between related states A and B can be computed from MD simulations by the use of the OSP method.[13] OSP is a direct application of the free energy perturbation (FEP) technique

by Zwanzig,[12a] and makes use of an (unphysical) reference state, R, from which free energy differences with respect to a set of real compounds, ΔG_{RA}, can be deduced, according to:

$$\Delta G_{RA} = G_A - G_R = -k_B T \ln\left\langle e^{-(H_A - H_R)/k_B T} \right\rangle_R \qquad \text{(Eq. 7.2)}$$

where k_B is the Boltzmann constant, T is the temperature, and H_A and H_R are the Hamiltonians pertaining to the real compound A and the reference state R, respectively. $\langle\ \rangle_R$ denotes ensemble averaging over all the configurations generated in a MD simulation using the Hamiltonian of state R. OSP makes use of two MD simulations of R, in aqueous and protein environment. The free energy difference between states A and another real ligand B can then be derived by using:

$$\Delta G_{AB} = \Delta G_{RB} - \Delta G_{RA} \qquad \text{(Eq. 7.3)}$$

Successful application of the OSP method requires the design of a reference molecule able to sufficiently sample the phase space relevant for the real states, *i.e.* the Boltzmann weight of the configurations sampled with Hamiltonian H_R should not vanish in terms of the Hamiltonian pertaining to the real compounds.

In this study, the representation of the reference state, K*R*K (depicted in Table 7.1), was designed such that sufficient overlap with the conformational space of real compounds, KGK and KAK was possible. *I.e.* the methyl group of R, Cβ, is treated as a soft atom with Van der Waals parameters of a CH3-group and a softness parameter α of 1.51. The original reference state, referred to as N, holds the same parameter set as real compound KAK.
Furthermore, in order to complement the alternative TI calculations which were carried out with a slightly modified Hamiltonian (described in Section 7.3.2), excluding specific intra-molecular interactions, an alternative OSP simulation was carried out with a slightly modified reference state, excluding the same intra-molecular interactions. This reference state is referred to as N+E.

7.2.3 Simulation settings

All molecular dynamics simulations were performed using the GROMOS11 biomolecular simulation package.[14-15] Interaction parameters to describe the tripeptide ligands and proteins according to pH 7 were taken from the GROMOS parameter set 45A4.[16]

All protein simulations were carried out using as initial structure one of the three crystal structures of OppA in complex with KGK (PDB code 1B3L),[5c] KAK (PDB code 1JET)[6] and

KSK (PDB code 1B51).[5c] The starting configurations of the tripeptides in the water simulations were directly extracted from their respective crystal structures.
The ligands and protein-ligand complexes were solvated in periodic rectangular boxes, containing ~2000 and ~15,500 SPC water molecules,[17] respectively, using a minimum solute to wall distance of 0.8 nm.

In all MD simulations, a time step of 2 fs was used. In an equilibration phase, the system was gradually heated up to 298 K, in three (water) and seven (protein) separate simulation steps of 20 ps each. In the first six protein equilibration steps, position restraints were applied on the protein and ligand atoms, using a gradually decreasing force constant (starting from 2.5 x 10^4 kJ mol^{-1} nm^{-2}). Covalent bonds were kept constrained using the SHAKE algorithm,[18] with a relative tolerance of 10^{-4}. During the production simulations, the temperature was kept constant (298 K) as well as the pressure (1 atm) by using the weak coupling algorithm,[19] with coupling times of 0.1 ps and 0.5 ps, respectively. The isothermal compressibility was set to 4.575 x 10^{-4} (kJ $mol^{-1}nm^{-3})^{-1}$. At every time step, nonbonded interactions were calculated up to a distance of 0.8 nm, by means of a pairlist that was generated every five steps. At every fifth time step, intermediate-range interactions at distances up to 1.4 nm were calculated and kept constant between updates. To account for electrostatic interactions beyond the cut-off radius of 1.4 nm, a reaction-field contribution was added to the energies and forces,[20] with a dielectric permittivity of 61.[21] During the simulations, coordinates were stored every 250 steps (0.5 ps) for the TI simulations at intermediate λ-points, and every 100 (0.2 ps) steps for the reference state simulations.

TI simulations were carried out between all pairs of the tripeptide ligands, depicted in Table 7.1, both in water environment and bound to the OppA protein. For one alchemical transformation, 11 equidistant λ-values were used. After an equilibration time of 20 ps, simulations of 1 ns were performed at every λ-value (unless stated differently), such that the total path of a transformation was 11 ns long. No position restraints were applied. In the water TI calculations, alternative protocols were carried out as well in order to allow for better convergence. Further details are described in Section 7.3.2.

For OSP, production MD simulations were performed for the reference states N and N+E at 298K without any position restraints, both in aqueous solution (20 ns), and bound to the OppA protein (10 ns).

7.3 Results and discussion

7.3.1 Thermodynamic integration

TI calculations were performed to compute the free energy differences between the three ligands, KGK, KAK and KSK free in solution and in complex with the protein. Initial TI calculations in protein environment to calculate free energy difference between KGK and KAK (corresponding to a very small chemical difference of a single methyl group) were performed using 100 ps per λ-point, leading to an overall simulation time of 1.1 ns per TI perturbation. Surprisingly, this minimal change resulted in a very poor hysteresis between the forward and backward simulations (data not shown). Initially, it was hypothesized that the TI simulations did not properly account for the different configuration of water molecules in the active site (see Figure 7.1, panels A and B). However, the same TI process (KAK into KGK) showed poor convergence when performed free in solution, as well. Based on this surprising finding, we decided to study the free energy difference in aqueous solution in more detail first.

7.3.2 TI free in solution

The values of $\langle \partial H/\partial\lambda \rangle_\lambda$ as a function of λ for the TI process KAK to KGK free in solution are shown in Figure 7.2. Depicted are the values for the forward and backward processes, shown in black and red solid lines, respectively. The corresponding numerically integrated values, resulting in the free energy difference between the two states, are presented in Table 7.2.

Initially, the TI protocol to alchemically transform KAK into KGK was performed using simulation times of 100 ps per λ-point. Using this protocol, however, poor convergence was obtained, as can be observed from the large hysteresis in Figure 7.2A, and the unconverged values in Table 7.2 for line A (-4.4 and 8.3 kJ mol^{-1}). This TI process lead to a hysteresis of 3.9 kJ mol^{-1} for the solvated compounds. These simulations were subsequently prolonged with 900 ps to 1 ns per λ-point, leading to the perturbation profiles shown in Figure 7.2B. Even though the hysteresis is reduced to 1.7 kJ mol^{-1} (Table 7.2), the forward and backward curves are still largely dissimilar.

It was observed that the tripeptide ligands are highly flexible, an effect which is illustrated in Figure 7.3. Here, the φ_2- and ψ_2-angles are monitored over 1 ns at every individual λ-point of the forward and backward perturbation process between KAK and KGK. In the left top corner, the system's Hamiltonian $H(\lambda)$ describes compound KAK, and in the right bottom corner compound KGK. The lower curves originate from the forward TI calculation while the upper curves in every panel originate from the backward calculations.

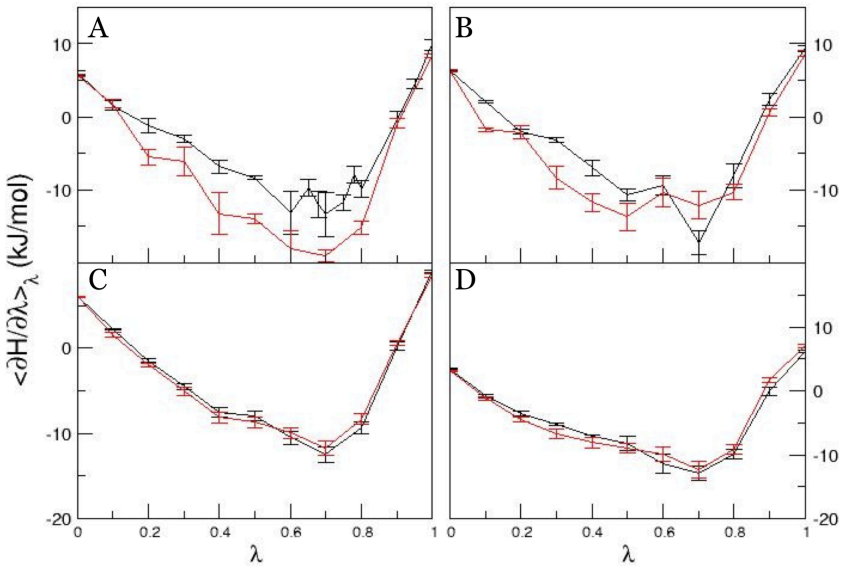


Figure 7.2. Values of $\langle \partial H/\partial \lambda \rangle_{\lambda}$ as a function of λ for the eight TI processes between KAK (λ = 0) and KGK (λ = 1) free in solution. Forward (black) and backward (red) perturbations were carried out in four TI protocols. For ease of comparison, the curves for the backward calculations are inverted using $\lambda \rightarrow 1\text{-}\lambda$ and $\langle \partial H/\partial \lambda \rangle \rightarrow -\langle \partial H/\partial \lambda \rangle$. In Panel A, 20 ps of equilibration time was followed by a production run of 100 ps per λ-point. In Panels B and C, the simulation time per λ-point was prolonged to 1 ns and 10 ns, respectively. In Panel D, production runs of 1 ns/λ-point were carried out, however with a slightly modified Hamiltonian, excluding specific intra-molecular interactions (see text for details).

TI		ΔG (kJmol^{-1})		
Panel Figure 7.2		*forward*	*backward*	*hysteresis*
A	100ps/λ	-4.4 ±1.0	8.3 ± 1.2	3.9
B	1ns/λ	-4.5 ± 0.8	6.2 ± 1.1	1.7
C	10ns/λ	-4.3 ± 0.4	4.4 ± 0.5	0.1
D	excl, 1ns/λ	-5.4 ± 0.6	5.4 ± 0.7	0.0
	excl, 1ns/λ + end point corrections	-3.6	4.2	0.6
	excl, 1ns/λ + reweighting	-4.5	4.5	0.0

Table 7.2. Calculated free energy differences ΔG (in kJ mol^{-1}) between hydrated KAK and KGK, as obtained from the various TI calculations depicted in Figure 7.2. See the caption of that figure for details on the calculations in the different panels and see the text for details on the corrections applied to the calculations in panel D.

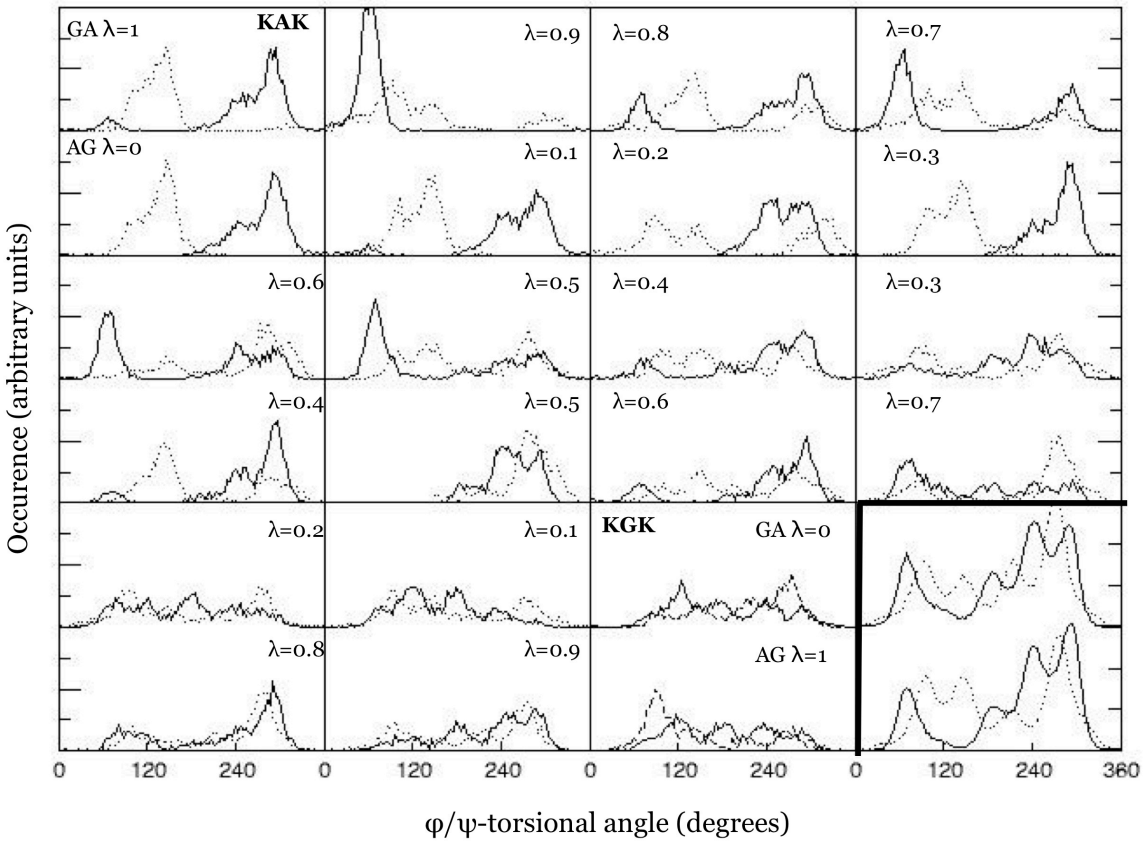


Figure 7.3. Distributions for the φ_2- and ψ_2-angles of residue X in ligand KXK, calculated for all λ-values (1ns) in the TI processes between KAK and KGK depicted in Figure 7.2B. The φ_2-distribution is represented by a black solid curve, the ψ_2-distribution by a dotted black curve. The distributions for the forward process are depicted in the lower half of the panel and those for the backward process in the upper half. Hence, distributions within one panel correspond to identical Hamiltonians. The distributions for the OSP simulations (20 ns) of reference states N and N+E in water are shown in the lower and upper half, respectively, of the last panel.

From the distributions in the end points it is clear that φ_2 and ψ_2 show distinct maxima in their distributions for KAK, while irregular, wide distributions are obtained for KGK. More importantly, the distributions at identical states from the forward and backward processes seem widely different at intermediate λ-values. This diverse sampling offers an explanation for the different estimates of $\langle \partial H/\partial \lambda \rangle$ at these λ-values.

Connecting the distributions in Figure 7.3, to the flexibility of the intermediate states of the ligand, it was necessary to modify the TI protocols. Prolongation to 10 ns per λ-point resulted in excellent convergence between forward and backward profiles (Figure 7.2C) and a corresponding hysteresis of 0.1 kJ mol^{-1} (Table 7.2). However, an overall simulation time of 110 ns per free energy difference remains impractical for systems only slightly larger than the ones presented here.

To circumvent the necessity of such large simulation times, a solution was found by slightly altering the Hamiltonian of the ligand in the TI calculations, leading to a smoother TI profile. For KAK and all intermediate states, the nonbonded interactions for atom CB_{19} with atoms C_{14}, H_{17}, O_{21} and N_{22} are excluded from the energy calculations in addition to the regular excluded atoms (1,3 neighbors). For atom numbering, see Table 7.1. These additional exclusions remove energy barriers for the transitions of φ_2 and ψ_2 and lead to the smooth TI profile in Figure 7.2D for simulation times of 1 ns per λ-point, in which no hysteresis is observed (Table 7.2).

However, the smooth profile for $\langle \partial H/\partial\lambda \rangle_\lambda$ in Figure 7.2D is obtained by an incorrect representation of the system's Hamiltonian. We applied two methods to correct for this.
In the first approach, the free energy estimate, $\Delta G_{TI,Excl}$ can be corrected by perturbing the system back to the correct Hamiltonian at both end states of the TI process. In order to do this, the OSP formula is applied as follows:

$$\Delta G_{X,corr} = G_{No\ Excl} - G_{Excl} = -k_B T \ln\left\langle e^{-(H_{NoExcl} - H_{Excl}/k_B T} \right\rangle_{X,Excl} \qquad \text{(Eq. 7.4)}$$

and the corrected free energy difference is obtained from:

$$\Delta G_{AG,\ No\ Excl} = \Delta G_{AG,\ Excl} + \Delta G_{KGK,\ corr} - \Delta G_{KAK,\ corr} \qquad \text{(Eq. 7.5)}$$

Using only 2000 snapshots for the application of Eq. (7.4) leads to an increase of the hysteresis to 0.6 kJ mol^{-1} following this approach (Table 7.2).

In the second approach, the entire free energy profile is corrected by considering the addition of the exclusions as a potential energy bias and unbiasing the estimate of $\langle \partial H/\partial\lambda \rangle$ like in the umbrella sampling approach.[22] The reweighted ensemble average for the correct Hamiltonian without the additional exclusions is obtained from:

$$\left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{NoExcl} = \frac{\left\langle \frac{\partial H}{\partial \lambda} e^{-(H_{NoExcl} - H_{Excl})/k_B T} \right\rangle_{Excl}}{\left\langle e^{-(H_{NoExcl} - H_{Excl})/k_B T} \right\rangle_{Excl}} \qquad \text{(Eq. 7.6)}$$

As can be seen in Table 7.2, this approach maintains a zero hysteresis and is in excellent agreement with the data for the 10 ns / λ-point simulations. In total, Figure 7.2 is based on 268 ns of MD simulations. It is therefore crucial to device more efficient protocols and approaches while maintaining the correct free energy differences.

In the TI perturbations for the (forward and backward) transformations KAK → KSK and KSK → KGK there were no convergence problems, and smooth profiles were obtained using simulation times of 1 ns per λ-point (data not shown). Apparently, the effect of the serine side-chain (in the transformation between KGK and KSK) reduces the problem with the flexibility of glycine, observed in transformations between KAK and KGK. The same exclusions were also added to the Hamiltonian for the KGK to KSK simulations, leading to virtually identical free energy values (data not shown).

7.3.3 TI bound to the protein

In a next step, TI calculations were performed within the protein complexes, according to the thermodynamic cycle depicted in Figure 7.4.

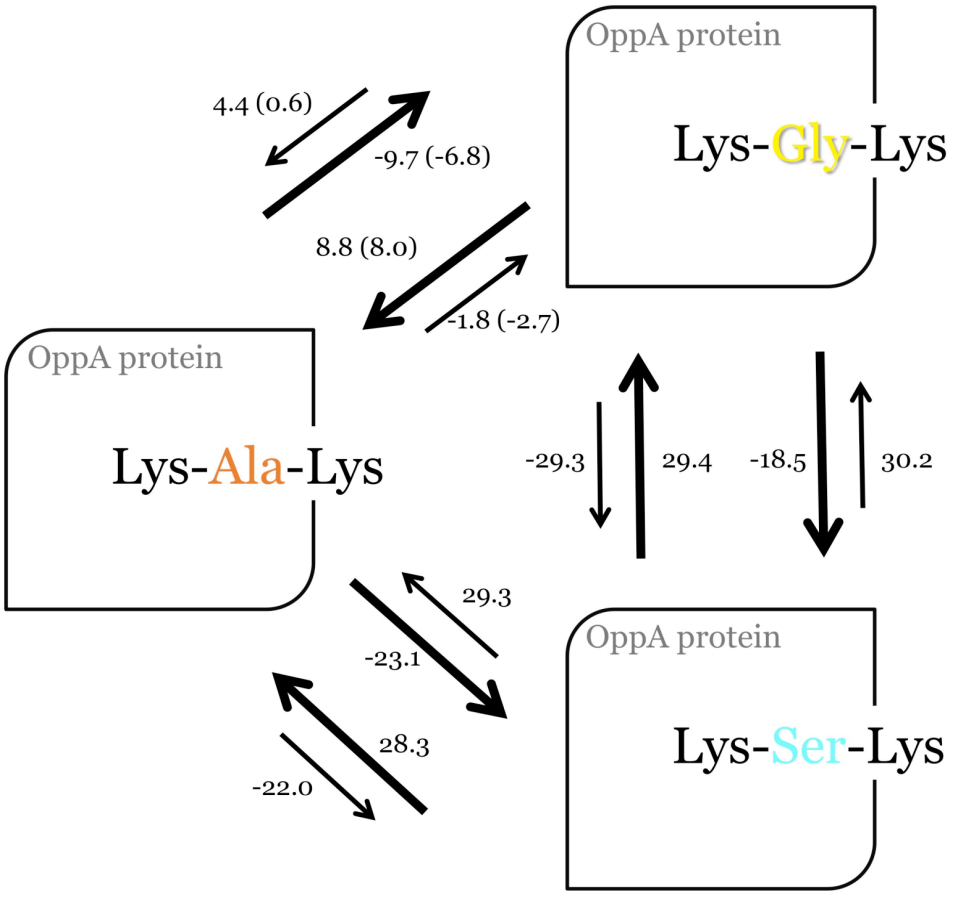


Figure 7.4. Thermodynamic cycle between three OppA ligands bound to the protein, obtained by TI between KGK, KAK and KSK. Relative free energies (ΔG) corresponding to the alchemical transformations are shown next to the arrows. Transformations between KAK and KGK have been performed twice, carried out with a slightly altered Hamiltonian, namely atoms C_{14}, H_{17}, O_{21} and N_{22} are taken up in the nonbonded exclusion list for CB_{19}; these values are shown in brackets. All TI calculations were carried out using two different starting configurations, originating from the crystal structures of the respective systems. The origin of the larger arrows indicates the crystal structure from which the simulations were started, while the smaller arrows represent the corresponding backward simulations.

Figure 7.1 shows distinct solvation patterns of the OppA active sites for the three respective crystal structures. More solvent molecules occupy the space between the glycine residue in KGK and the OppA protein (panel A), than in the other two complexes. Panels B and C show less solvent molecules at that location, since the side chains of alanine (KAK) and serine (KSK) residues occupy the space. In order to properly consider these distinct solvation patterns, two independent pairs of TI perturbations were performed per transformation, each starting from different starting configurations, originating from the respective crystal structure of the tripeptide-OppA complexes. The origin of the thicker arrows in Figure 7.4 represent the starting X-ray structures of the TI simulations. The smaller arrows represent the corresponding backward simulations. The accompanying forward and backward values of the relative free energies between the tripeptides calculated from the various simulations are indicated next to their respective arrows in Figure 7.4 as well.
Since initial TI calculations, with a simulation time of 100 ps per λ-point, showed very poor convergence, all TI calculations depicted in Figure 7.4 are computed using simulation times of 1 ns per λ-point. This adds up to 11 ns per single TI run. Overall, Figure 7.4 depicts a total of 3 x 4 x 11 = 132 ns simulation time.

The obtained values of $\langle \partial H/\partial \lambda \rangle_\lambda$ as a function of λ for all the TI processes (forward and backward) in Figure 7.4 are shown in Figure 7.5.

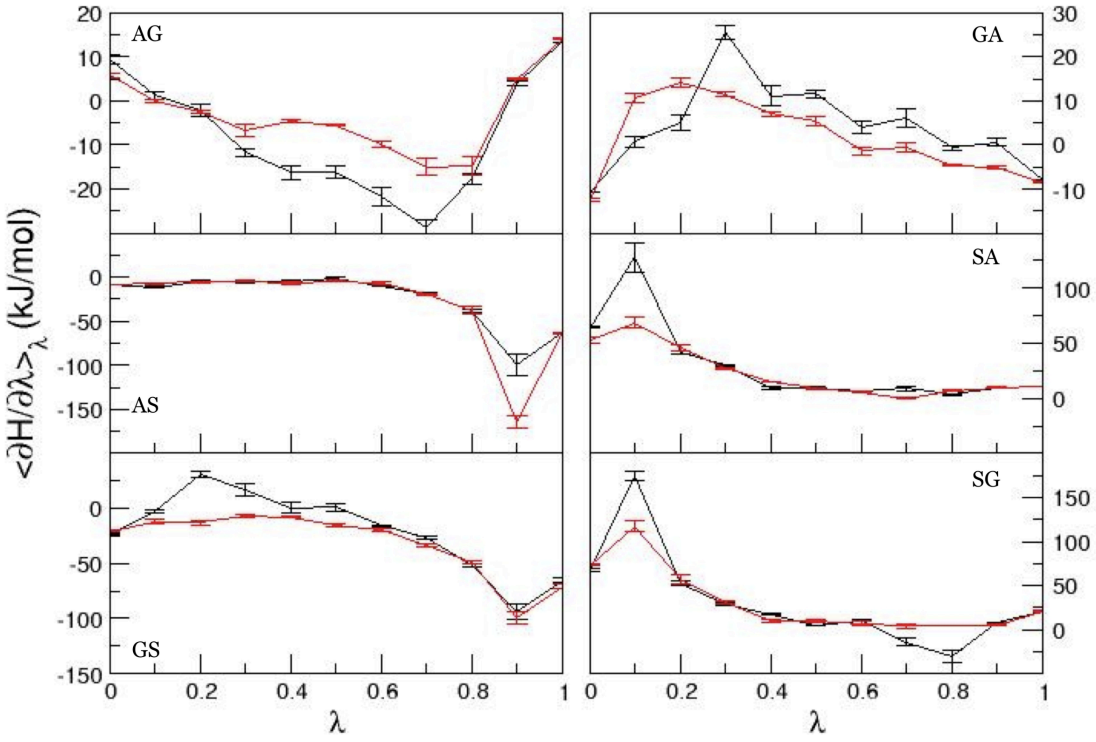


Figure 7.5. Values of $\langle \partial H/\partial \lambda \rangle_\lambda$ as a function of λ for the TI perturbation processes shown in the thermodynamic cycle in Figure 7.4. Black curves correspond to forward, red curves to backward simulations. For ease of comparison, backward curves are inverted using $\lambda \rightarrow 1\text{-}\lambda$ and $\langle \partial H/\partial \lambda \rangle \rightarrow -\langle \partial H/\partial \lambda \rangle$.

Even for cases in which the corresponding integrated values, shown in the thermodynamic cycle (Figure 7.4) seem to show low hysteresis, Figure 7.5 reveals that the actual TI profiles do not agree. *E.g.*, whereas KSK → KGK and backwards amounts to converged values in Figure 7.4 (29.4 and -29.3 kJ mol^{-1}, respectively), the corresponding $\langle \partial H/\partial\lambda \rangle_\lambda$ profiles depicted in Figure 7.5 show discrepancies at λ=0.1; λ=0.7 and λ=0.8. Most probably, a cancellation of errors has occurred leading to (seemingly) converged forward and backward values for ΔG.

In spite of the remaining hysteresis, the average of all TI processes corresponding to a given transformation, free in solution and bound to the protein, is calculated in Table 7.3. From these values, relative binding free energies are calculated and compared to the experimental values.

The TI prediction for GS (KGK → KSK) is in excellent agreement with experiment, with an error of merely 0.2 kJ mol^{-1}. However, in Figure 7.5 it is shown that KGK → KSK and KSK → KGK show meager convergence between the forward and backward TI profiles. These discrepancies are also reflected by a cycle closure of -7.5 kJ mol^{-1} for the protein environment. For the free states, the thermodynamic cycle closes with -1.2 kJ mol^{-1} (Table 7.3).
The calculated relative binding free energies of KAK and KGK (AG) do not agree with the experiment. While the calculated value for $\Delta\Delta G_{bind}$ amounts to -1.7 kJ mol^{-1} in favor of KGK, $\Delta\Delta G_{bind,exp}$ is 7.5 kJ mol^{-1} in favor of KAK. A smaller discrepancy of 3.1 kJ mol^{-1} is observed between KAK and KSK.

TI perturbation	Free	Protein	$\Delta\Delta G_{bind}$	$\Delta\Delta G_{bind,exp}$
AG	-4.5	-6.2	-1.7	7.5
GS	-18.3	-26.9	-8.6	-8.4
AS	-21.6	-25.6	-4.0	-0.9
cycle	*-1.2*	*-7.5*	*-6.3*	*0*

Table 7.3. Comparison of experimentally determined $\Delta\Delta G_{bind,exp}$ for all tripeptides bound to the OppA protein, to values for $\Delta\Delta G_{bind}$ calculated from TI simulations in free and bound states. For each transformation, the forward and backward values are averaged, as are the independent protein values in Figure 7.4. The values for the overall cycle are obtained as $\Delta G_{AG} + \Delta G_{GS} - \Delta G_{AS}$ and should theoretically amount to zero. All free energy differences are given in kJ mol^{-1}.

The TI between compounds KAK and KGK in complex with OppA was also carried out with the previously described modifications to the Hamiltonian, excluding specific intra-molecular interactions. Atoms C_{14}, H_{17}, O_{21} and N_{22} are taken up in the nonbonded exclusion list for CB_{19}. The values for ΔG corresponding to the various TI perturbations using extra exclusions, are shown in Figure 7.4 in brackets.
Whereas these extra exclusions significantly improved the poor convergence of the TI profiles for the conversion of KAK → KGK in water (Figure 7.2D), in protein this was not the case. Apparently, the flexibility of the glycine residue in KGK is not (exclusively) the cause of the large hysteresis anymore, which is not surprising considering the intrinsically reduced flexibility in the protein.

Overall, the calculated relative binding free energies of the compounds KAK, KGK and KSK bound to the OppA protein were in poor agreement with experimentally determined values.

7.3.4 Densities of water molecules in the active site

The poor convergence of the obtained TI values for KAK, KGK and KSK (displayed by the hystereses in Figure 7.5) are likely arising from inaccurate solvation of the active site during the TI process (crucial to the binding thermodynamics of OppA).

This is illustrated in Figure 7.6, which shows four end points of the TI processes $KGK_{X\text{-}ray}$ → KSK. As can be seen in Figure 7.4, this transformation displays one of the largest hystereses, amounting to 11.7 kJ mol^{-1}. In panels A and D, the system's Hamiltonian $H(\lambda)$ describes compound KGK in complex to the OppA protein, and in panels B and C, compound KSK. As shown in the Xray structures depicted in Figure 7.1, more crystallographic water molecules are present in the KGK active site (panel A). In KSK (Figure 7.1B), these interfacial water molecules are displaced by the serine side chain. In the forward perturbation process, $KGK_{X\text{-}ray}$ → KSK, the water molecules indeed seem to be replaced by the serine side chain, resulting in a lower water density in panel B than in panel A. However, in the backward process, the water density is not being restored. Comparing panel D to panel A, it becomes clear that the water molecules do not fill the gap between the glycine and the protein after the backward simulation.

A closer investigation of panels B and C in Figure 7.6 may yield additional insight. The densities in these panels correspond to the same state of the system and were obtained from two consecutive simulations. Yet, the density in panel B is still more pronounced than in panel C, suggesting that the dynamics of the water molecules has a slower relaxation time than 1 ns. In fact, comparing panel D to panel C suggests that the water density appropriate for KGK is being restored, but that more time will be required to reach the densities observed in panel A.

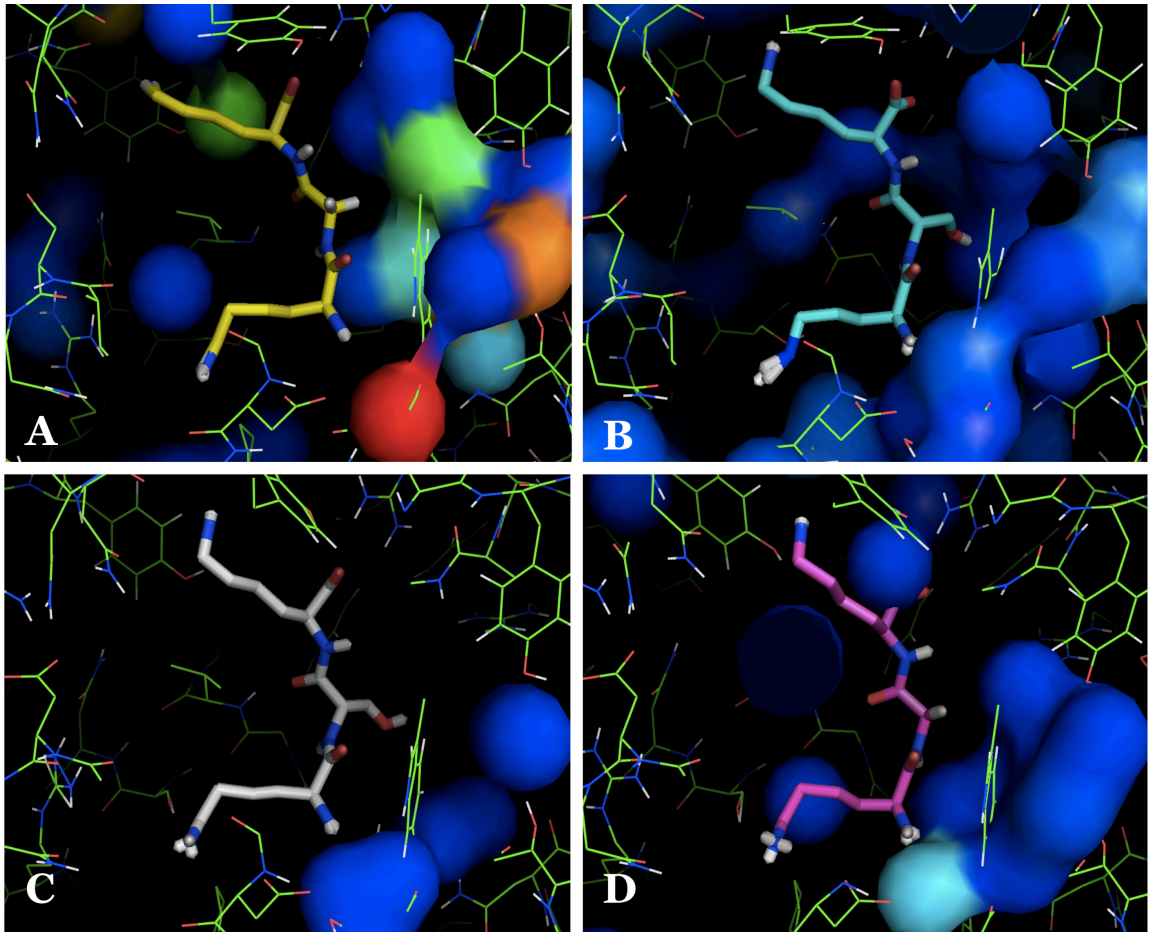


Figure 7.6. Depiction of average water densities over the end-point TI simulations for KGK to KSK, originating from the crystal structure of KGK in complex with OppA. Panel A: λ=0 of the forward KGK to KSK simulation; panel B: λ=1 of the same simulation; Panel C: λ=0 of the backward KSK to KGK simulation; panel D: λ=1 of the same simulation. Panels A & D and B & C represent identical states. Protein and ligand atoms are depicted at their average positions after a superposition based on the backbone atoms. Due to fast rotations the depiction of some of the terminal groups seem distorted. The average water density was monitored on a grid with spacing of 0.2 nm. Water molecules are assigned to their nearest grid point. Grid points with at least 30% occupancy of the highest observed occupancy are indicated in the figure and colored accordingly (red ~90%; green ~50%; dark blue ~30%). For clarity, some water locations close to the lysine side chains of the ligand have been removed.

This analysis suggests that longer simulations may overcome this sampling problem. However, it should be noted that this will be necessary for all intermediate λ-values.

A similar slow convergence may be causing the hysteresis in the KAK → KSK transitions in Figure 7.4. It is striking that the two KAK → KSK transformations, although originating from different X-ray structures, seem to agree closely and that the same is true for the two KSK → KAK transformations, *i.e.* the hysteresis is very reproducible, and does not depend on which simulation is performed first. This suggests that the transformation processes follow similar pathways, but are not sufficiently long to rearrange the solvent network sufficiently.

7.3.5 One-step perturbation

To complement the TI results, OSP calculations were carried out for two tripeptide ligands binding to OppA, namely KAK and KGK. Two different reference states (N and N+E) were used in simulations in water and protein environments. The free energy differences of KAK and KGK with respect to the reference state were calculated using OSP and are shown in Table 7.4. The relative binding free energies were also calculated based on these values and are compared to the experimentally measured difference in binding affinity of KGK and KAK.
With regard to Table 7.4, two distinct observations are to be made.

OSP					**Experiment**
		Free	Protein	ΔG_{bind}	$\Delta G_{bind,\,exp}$
Reference state used					
	$\Delta G_{REF \rightarrow A}$	6.4	3.3	-3.1	-41.1
N	$\Delta G_{REF \rightarrow G}$	2.1	13.3	11.2	-33.6
	$\Delta\Delta G_{bind}$			*-14.3*	*-7.5*
	$\Delta G_{REF \rightarrow A}$	4.4	4.1	-0.3	-41.1
N + E	$\Delta G_{REF \rightarrow G}$	-0.5	10.3	10.8	-33.6
	$\Delta\Delta G_{bind}$			*-11.1*	*-7.5*

Table 7.4. Comparison of experimentally determined $\Delta\Delta G_{bind,exp}$ for KAK and KGK bound to the OppA protein, to values obtained from the differences in free energy ΔG between reference and end states, calculated from water (free) and protein ensembles generated in the one-step perturbation approach using the N and N+E reference states. Water and protein MD simulations were 20 ns and 10 ns long, respectively. All free energy differences are given in kJ mol^{-1}.

Firstly, the binding free energies calculated using OSP agree significantly better with the experimental data than the values calculated using TI. Not only is the preference of OppA to KAK rather than KGK reproduced (in contrast to the TI data, see Table 7.3), the calculated binding affinities also quantitatively agree reasonably with experiment. This is particularly the case in the OSP calculations using reference state N+E, where the calculated value of -11.1 kJ mol^{-1} is to be compared to the experimentally determined value of -7.5 kJ mol^{-1}, with a discrepancy of 3.6 kJ mol^{-1}.

The second observation to be made, is that the results for ΔG_{bind}, obtained by using the newly defined reference state N+E improved with 3.2 kJ mol^{-1}, decreasing from -14.3 kJ mol^{-1} to -11.1 kJ mol^{-1} in comparison to the original reference state N. This is mostly due to the calculation in

the protein environment. From Table 7.4 estimates can also be obtained for ΔG_{AG}. Free in solution values of -4.3 kJ mol^{-1} (N) and -4.9 kJ mol^{-1} (N+E) are obtained, in good agreement with the converged values in Table 7.2, (-4.5 kJ mol^{-1}, line D). In the protein environment, values of ΔG_{AG} = 10 kJ mol^{-1} (N) and 6.2 kJ mol^{-1} (N+E) are obtained, which fluctuate more and are in stark contrast to the TI data in Figure 7.4. The good agreement with the TI data for the tripeptides free in solution most likely comes from the fact that both reference states sufficiently sample the φ_2- and ψ_2-angles of the real tripetides (lower right panel in Figure 7.3).

It seems that OSP yields better results for this particular system than TI calculations. Not only is the internal ranking of binding between KAK and KGK correct, the calculated values are in fair agreement to experimentally measured affinities.

When reviewing the results for calculated free energy values using TI and OSP, we show once more (in addition to Chapter 4) that TI is not necessarily the most reliable method, but that similar or even more accurate results can be obtained when applying the less expensive OSP method on the very same system. However, in this study, only a minimal alchemical change was pursued; the transformation of Ala to Gly is very small. However, this already appeared to be a challenge for the most straightforward TI setup, showing large hystereses in the solvated state. This challenge could be overcome by altering the system's Hamiltonian, applying more exclusions in the nonbonded term for certain atoms. However, while this modification could solve the large discrepancy in the solvated environment, the TI profiles obtained for the protein complexes remained unconverged.

The results described in this chapter emphasize once more that OSP offers an elegant way to calculate free energy differences between (similar) compounds, and in comparison to TI may be more efficient and even more accurate. This suggests that OSP may very well be a reliable method to calculate free energies. However, to increase its robustness, once should study OSP further to explore its limits, *e.g.* imposed by a limited chemical diversity of the real compounds.

In OSP, the reference state should be chosen such that sufficient overlap is possible with the real compounds, in this case the KAK and KGK end states. Evidently, a next challenge would be to extend the set of real compounds to more KXK ligands, *e.g.* KSK. However, a representation of the reference state such that the sampling is representative for both Gly and Ser is much more challenging.[23]
Not only would it require two heavy atoms to be grown in or vanish, a water molecule needs to be occasionally pushed away and configurations need to be sampled that are representative of a polar OH-group, potentially forming H-bonds with its surroundings. This would probably require the careful construction of a reference state, as described in Chapter 5. Currently, this remains out of reach and therefore it was not incorporated in this study.

7.4 Conclusion

Much valuabe information on the thermodynamic properties and structural features of oligopeptides binding to OppA is available, shedding light on the molecular recognition process. However, exploring the (complicated) binding behavior of the flexible tripeptide ligands to OppA using free energy methods, based on statistical thermodynamics, appears to be a challenge. First of all, the high flexibility of the glycine residue had a considerable effect on the convergence of the TI calculations in the water environment. By excluding specific intra-molecular interactions, this problem could be overcome. However, when this modified Hamiltonian was applied to the system in the bound environment, problems with internal consistency of the simulated data remained. Even the reasonable agreement with experimental affinities for the KGK/KSK and KAK/KSK pairs should not be trusted. Apparently, the flexibility of the glycine residue is not causing large discrepancies here, but rather the water molecules that were shown to display different local densities in forward and backward TI processes.
For the smallest chemical modification, KGK → KAK the OSP method offers an elegant solution, with reasonable agreement between calculated and experimental binding free energies.
For the larger alchemical changes a more elaborate reference state may be devised, or significantly longer simulations may be required to capture the slow dynamics of the active site water molecules.

We can conclude that a full thermodynamic characterization using free energy calculations of the OppA system remains demanding, considering the flexibility of the ligands and the variable solvation patterns.

References

(1) (a) Levitt, M.; Park, B. H. Water: now you see it, now you don't. *Structure* **1993,** *1*, 223-226; (b) Li, Z.; Lazaridis, T. The effect of water displacement on binding thermodynamics: Concanavalin A. *J. Phys. Chem. B* **2005,** *109*, 662-670; (c) Li, Z.; Lazaridis, T. Water at biomolecular binding interfaces. *Phys. Chem. Chem. Phys.* **2007,** *9*, 573-581.

(2) Monnet, V. Bacterial oligopeptide-binding proteins. *Cell Mol. Life Sci.* **2003,** *60*, 2100-2114.

(3) (a) Guyer, C. A.; Morgan, D. G.; Staros, J. V. Binding specificity of the periplasmic oligopeptide-binding protein from *Escherichia coli*. *J. Bacteriol.* **1986,** *168*, 775-779; (b) Goodell, E. W.; Higgins, C. F. Uptake of cell wall peptides by Salmonella typhimurium and *Escherichia coli*. *J. Bacteriol.* **1987,** *169*, 3861-3865.

(4) (a) Hammond, S. M.; Claesson, A.; Jansson, A. M.; Larsson, L. G.; Pring, B. G.; Town, C. M.; B., E. A new class of synthetic antibacterials acting on lipopolysaccharide biosynthesis. *Nature* **1987,** *327*, 730-732; (b) Kornacki, J. A.; Oliver, D. B. Lyme disease-causing Borrelia species encode multiple lipoproteins homologous to peptide-binding proteins of ABC-type transporters. *Infect. Immun.* **1998,** *66*, 4115-4122.

(5) (a) Tame, J. R. H.; Dodson, E. J.; Murshudov, G.; Higgins, C. F.; Wilkinson, A. J. The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands. *Structure* **1995,** *3*, 1395-1406; (b) Davies, T. G.; Hubbard, R. E.; Tame, J. R. H. Relating structure to thermodynamics: the crystal structures and binding affinity of eight OppA-peptide complexes. *Protein Sci.* **1999,** *8*, 1432-1444; (c) Sleigh, S. H.; Seavers, P. R.; Wilkinson, A. J.; Ladbury, J. E.; Tame, J. R. H. Crystallographic and calorimetric analysis of peptide binding to OppA protein. *J. Mol. Biol.* **1999,** *291*, 393-415.

(6) Tame, J. R. H.; Sleigh, S. H.; Wilkinson, A. J.; Ladbury, J. E. The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. *Nat. Struct. Biol.* **1996,** *3*, 998-1001.

(7) Ladbury, J. E. Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.* **1996,** *3*, 973-980.

(8) Wang, T.; Wade, R. C. Comparative binding energy (COMBINE) analysis of OppA-peptide complexes to relate structure to binding thermodynamics. *J. Med. Chem.* **2002,** *45*, 4828-4837.

(9) Tian, F.; Yang, L.; Lv, F.; Luo, X.; Pan, Y. Why OppA protein can bind sequence-independent peptides? A combination of QM/MM, PB/SA, and stucture-based QSAR analyses. *Amino Acids* **2011,** *40* (493-503).

(10) Berntsson, R. P.-A.; Doeven, M. K.; Fusetti, F.; Duurkens, R. H.; Sengupta, D.; Marrink, S.-J.; Thunissen, A.-M. W. H.; Poolman, B.; Slotboom, D.-J. The structural basis for peptide selection by the transport receptor OppA. *The EMBO Journal* **2009,** *28*, 1332-1340.

(11) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935,** *3*, 300-313.

(12) (a) Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* **1954,** *22*, 1420-1426; (b) Beveridge, D. L.; DiCapua, F. M. Free energy via molecular simulation: Applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Bio.* **1989,** *18*, 431-492.

(13) (a) Liu, H.; Mark, A. E.; van Gunsteren, W. F. Estimating the relative free energy of different molecular states with respect to a single reference state. *J. Phys. Chem.* **1996,** *100*, 9485-9494; (b) Schäfer, H.; van Gunsteren, W. F.; Mark, A. E. Estimating relative free energies from a single ensemble: Hydration free energies. *J. Comp. Chem.* **1999,** *20*, 1604-1617.

(14) Riniker, S.; Christ, C. D.; Hansen, H. S.; Huenenberger, P. H.; Oostenbrink, C.; Steiner, D.; van Gunsteren, W. F. Calculation of relative free energies for ligand-protein binding, solvation, and conformational transitions using the GROMOS software. *J. Phys. Chem. B* **2011,** *115*, 13570-13577.

(15) Schmid, N.; Christ, C. D.; Christen, M.; Eichenberger, A. P.; van Gunsteren, W. F. Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation. *Comp. Phys. Commun.* **2012,** *183*, 890-903.

(16) Lins, R. D.; Hünenberger, P. H. A new GROMOS force field for hexopyranose-based carbohydrates. *J. Comp. Chem.* **2005,** *26*, 1400-1412.

(17) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J., Interaction models for water in relation to protein hydration. In *Intermolecular Forces*, Pullman, B., Ed. Reidel: Dordrecht, The Netherlands, 1981; pp 331-342.

(18) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comp. Phys.* **1977,** *23*, 327-341.

(19) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **1984,** *81*, 3684-3690.

(20) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. A generalized reaction field method for molecular-dynamics simulations. *J. Chem. Phys.* **1995,** *102*, 5451-5459.

(21) Heinz, T. N.; van Gunsteren, W. F.; Huenenberger, P. H. Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations. *J. Chem. Phys.* **2001,** *115*, 1125-1137.

(22) (a) Torrie, G.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comp. Phys.* **1977,** *23*, 187-199; (b) König, G.; Boresch, S. Non-Boltzmann sampling and Bennett's acceptance ratio method: how to profit from bending the rules. *J. Comput. Chem.* **2011,** *32*, 1082-1090.

(23) Oostenbrink, C.; van Gunsteren, W. F. Single-step perturbations to calculate free energy differences from unphysical reference states: limits on size, flexibility, and character. *J. Comput. Chem.* **2003,** *24*, 1730-1739.